AN INDEX OF INCOMES

By K. G. KARSTEN, *New York City*

The purpose of this article is to suggest a set of two index figures or coefficients for the measurement and analysis of income and wealth statistics, which shall, without loss of statistical accuracy, be suitable for popular presentation. One of these coefficients gives a quantitative measurement and may be called an index of the size of incomes; the other is qualitative and may be called an index of the distribution of wealth. Both indices involve what may be called a natural law of incomes, with which, since the work of Pareto,* many statisticians are already acquainted, and of which a short explanation is prefixed.

When statistics of income are plotted upon double-logarithmic paper, that is, upon chart-paper ruled logarithmically on both axes, the resulting curve approaches a straight-line. Before plotting, the data must be cumulated downward, as there are no small frequency intervals of equal geometric distance;† and in this article wherever the figures of income are referred to, it is to be understood that the cumulated and not the individual class figures are meant. The cumulated figures (Tables I and II), representing the number of persons report-

---

* Vilfredo Pareto, *Cours d'Economie Politique*, Vol. II, pp. 304 *et. seq.*

The present writer unfortunately could not claim acquaintance with the work already done in this field at the time of his investigation and has, therefore, departed slightly from the method of Pareto, without, it must be said, great advantage in such departures. The use of fixation points selected nearer the central portions of the data, has, it is believed, some advantage in the avoidance of possibly artificial deviations at the lower extreme and of chance variations in the small number of quotations at the upper income extreme. The use of fixation points at income levels in the ratio of 10: 1 has the mathematical advantage of the short-cut method noted in the footnote on p. 259.

For graphic presentation, the present writer has finally projected the income levels along the vertical axis, as more readily connoting higher and lower incomes, in spite of the fact that this method projects the unknown variable along the horizontal axis. The so-called "factor of slope" (S), from which the index of distribution is derived, is, like the "inclination" used by Pareto, the tangent of the curve with the horizontal axis, but owing to this reversal of axes, is actually the reciprocal of Pareto's measure, that is, it is the co-tangent of the angle measured by him.

† Income statistics need not be cumulated to give the straight logarithmic curve, but the difficulty of showing the great number of small incomes separately makes cumulation convenient. If the statistics are not cumulated, the individual incomes must be thought of as arranged along a base line, or axis, with the largest incomes at one end and the smallest at the other, the remainder ranging themselves between, in succession. The position along the base line or axis must be on the logarithm of the successive number of the incomes; thus the largest income being at the beginning of the base line, the second largest must be at the logarithm of two, the third largest at the logarithm of three, the fourth largest at the logarithm of four, . . . to the smallest income at the logarithm of its successive number (around 100,000,000 in the United States).

All of this is implied in the cumulated figures used in the text; thus, if we have one person with an income of $10,000,000 and two with an income of $5,000,000 or more, the first being already indicated at $10,000,000, the second person has $5,000,000; if we have three persons with $3,333,333 or more, the first two being known, the third has $3,333,333; if we have four persons with $2,500,000 or more, the fourth one has $2,500,000; if we have five with $2,000,000 or more, the fifth has two millions, etc., etc.

TABLE I

CUMULATED INCOME STATISTICS SHOWING NUMBER OF PERSONS WITH INCOMES GREATER THAN EACH INCOME LEVEL. UNITED STATES, 1914

Income Classes (Incomes Between:)	Number of Persons	Cumulation (Incomes of More Than:)	Number of Persons
$1,000,000 and over	60	$1,000,000	60
500,000 to $1,000,000	114	500,000	174
400,000 to 500,000	69	400,000	243
300,000 to 400,000	147	300,000	390
250,000 to 300,000	130	250,000	520
200,000 to 250,000	233	200,000	753
150,000 to 200,000	406	150,000	1,159
100,000 to 150,000	1,189	100,000	2,348
50,000 to 100,000	5,161	50,000	7,509
40,000 to 50,000	3,185	40,000	10,694
30,000 to 40,000	6,008	30,000	16,702
25,000 to 30,000	5,483	25,000	22,185
20,000 to 25,000	8,672	20,000	30,857
15,000 to 20,000	15,790	15,000	46,647
10,000 to 15,000	34,141	10,000	80,788
5,000 to 10,000	127,448	5,000	208,236
4,000 to 5,000	66,525	4,000	274,763
3,000 to 4,000	82,754	3,000	357,515
Total (all classes)	357,515	Total (over $3,000)	357,515

TABLE II

CUMULATED INCOME STATISTICS.   UNITED STATES, 1914–1917, WITH PROFESSOR W. I. KING'S ESTIMATE FOR 1910

| Income Exceeded	King's Estimate (1910)	Number of Persons Having Greater Incomes			
		From Tax Reports			
		1914	1915	1916	1917
$10,000,000	5	....	....	....	....
5,000,000	15	....	....	10	4
4,000,000	....	....	....	19	12
3,000,000	....	....	....	33	17
2,000,000	56	....	....	67	41
1,500,000	....	....	....	109	74
1,000,000	154	60	120	206	141
500,000	....	174	329	582	456
400,000	....	243	451	827	635
300,000	....	390	705	1,296	1,015
200,000	415	753	1,307	2,449	2,060
150,000	....	1,159	2,031	3,733	3,362
100,000	3,560	2,348	3,824	6,633	6,664
75,000	....	....	5,880	....	....
50,000	15,190	7,509	10,671	17,085	19,103
40,000	22,190	10,694	14,771	22,696	26,190
30,000	34,190	16,702	21,776	32,764	38,923
20,000	54,190	30,857	37,679	53,775	66,300
15,000	....	46,647	54,154	76,396	96,196
10,000	127,190	80,788	88,256	121,701	161,996
8,000	165,190	....	....	157,309	....
6,000	244,190	....	....	....	....
5,000	....	208,236	208,658	272,252	432,662
4,000	449,190	274,763	267,607	344,279	618,467
3,000	720,190	357,515	336,652	429,401	993,425
2,000	1,437,190	....	....	....	1,832,132
1,500	2,702,190	....	....	....	....
1,000	8,543,190	....	....	....	3,472,890
800	13,542,190	....	....	....	....
600	20,657,190	....	....	....	....
400	25,942,190	....	....	....	....
300	27,655,190	....	....	....	....
200	27,925,190	....	....	....	....
100	27,945,190	....	....	....	....

NUMBER OF PERSONS:—
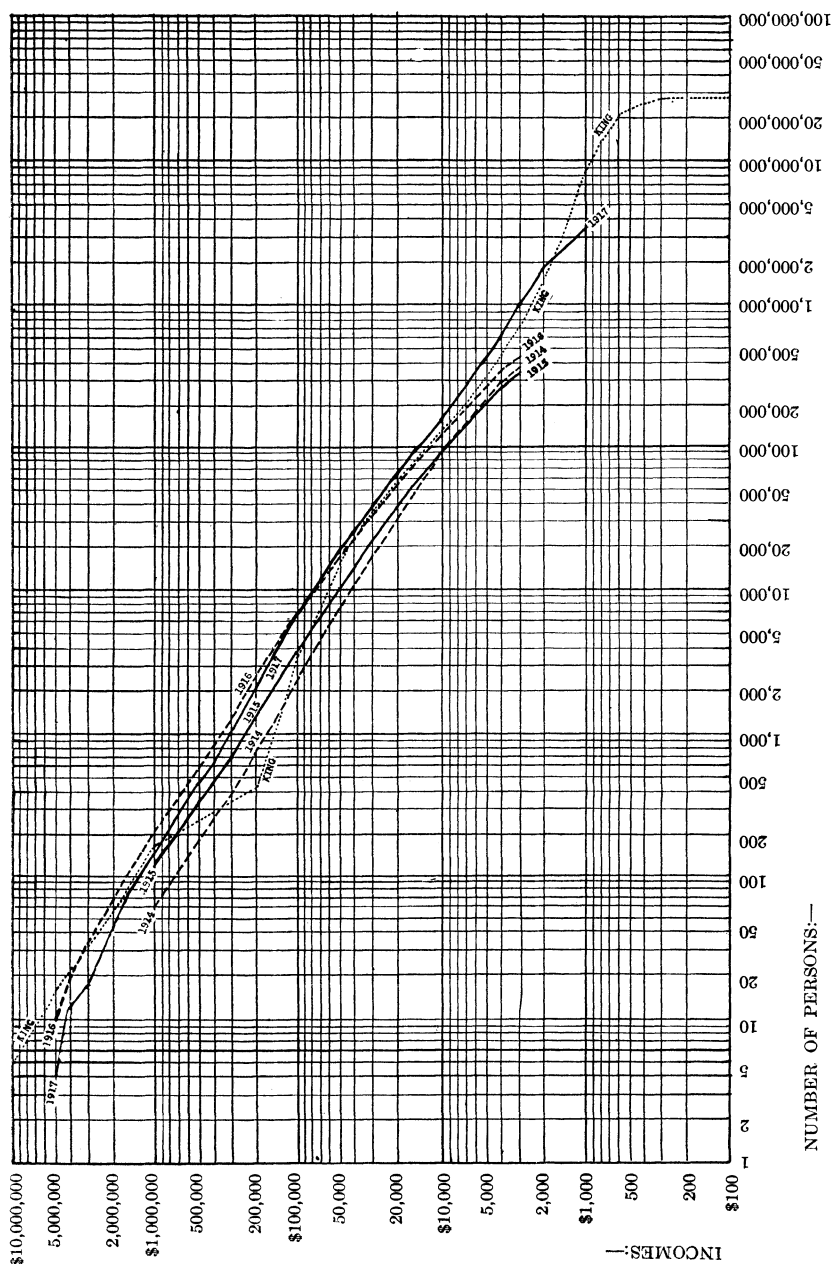
CHART I. GEOMETRIC OGIVE CURVES OF INCOMES
United States, 1910–17

Showing the ogive curves for the various years and for Professor King's estimate (dotted line), plotted upon logarithmic paper, in which the ordinates represent incomes in dollars and the abscissae represent the number of persons reporting more than such incomes.

ing incomes greater than each income-level, will when plotted (Chart I) form an ogive frequency curve, which may, since both axes are projected logarithmically, be called a geometric ogive curve.

The significance of the curves in the years actually reported (1914–17) lies in their very close approach to a straight line,* and it may, therefore, not be amiss to discuss for a moment their deviation from a straight-line (see Chart II); for it may be that this deviation is not
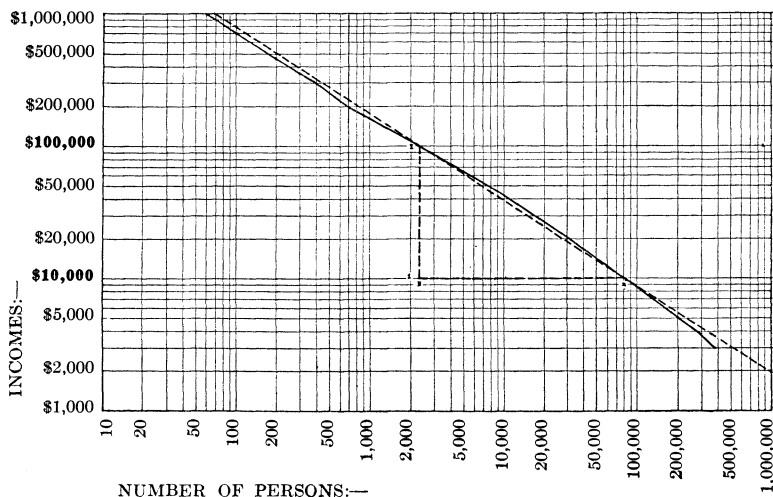

CHART II. NORM PROJECTED THROUGH FIXED POINTS
United States, 1914

Showing the slight deviation of the reported figures (full line), from the norm (dotted line) projected through the fixed points (circles) selected at $100,000 and $10,000. The values from which the index of distribution is computed are shown by the triangle the two legs of which are I/i and n/N.

real, but is due to faults in the statistics alone. It is stated in treasury reports that there has been considerable tax evasion, either wilful or negligent. Now it is easily seen that the larger the income, the greater is the temptation (especially under a graduated tax) to conceal portions of income. Were tax collectors humanly corruptible, it might be added that the facilities for concealment would also be greater. The temptation alone may account for a gradual deviation of the curve in the higher income levels. In the very small incomes, on the other hand, there is probably the greatest amount of negligent and unconscious

* Measured in terms of the tangent of the angle formed by the geometric ogive curve and the ordinates of the paper, the maximum angular deviation of the reported incomes curve from the straight line does not exceed 20 per cent, except at the lower tax limit.

evasion.   This includes the farmer who consumes part of the produce
of his own farm and makes no report on the produce so used, because
no money exchange has taken place to make this a money income.*
At the extreme lower tax limit, an abrupt drop in every curve is due
to the failure to make any tax report of persons whose incomes are
taxable by a small margin only.   Operating as these forces do, at differ-
ent ends, they may be expected to make the curve slightly convex—a
form it invariably shows—and the view may be held that the deviations
are not true of the phenomena, but are due to the statistical defects.

Obviously, if the true curve be a straight line, it would signify an
all-pervasive force or natural law, arranging in strict geometric order
or progression the fortunes and possessions of men, and bringing out
of the apparent chaos of human society a symmetry and simplicity
which is indeed fascinating.   This natural law might be phrased as
follows: "In any community the number of persons whose incomes
exceed any given amount varies inversely and logarithmically with
the given amount, modified by a constant factor peculiar to the com-
munity."   In the constant factor peculiar to the community we should
have a coefficient for the community, and although we knew the in-
come distribution through only a small range of incomes, we should
be able to unlock the secret of all other incomes in this community.

It is not necessary, however, to adopt the theory that the true curve
is a straight line in order to arrive at two useful and significant indices
or coefficients for the statistical analysis and comparison of income
conditions in various communities.   From the simple expedient of a
fitted straight-line, fitted *at standard and invariable points*, we can de-
rive two such coefficients of ample value in sociological analysis to
recommend them in the place of the present lack of relative wealth
indices.   For although the coefficients will not have definite individual
accuracy in themselves alone, yet the variations in these coefficients
for different communities, or for the same community at different times,
tell a significant story of change in much the same way as relative price
figures, derived from a large number of price quotations, tell a story of
price changes.

For the sake of simplicity, in this paper two coefficients are suggested
and their significance described as though the theory of the straight-
line were already an established fact.   The attempt is made to insert,
from time to time, warnings to the effect that this significance is still
theoretical.   Even if the curve finally prove to be (when the facts

* Some estimates of the value of produce consumed on the farm compiled from investigation among
Wisconsin farmers indicate that as much as 20 per cent of the total expense of the farm is for produce so
used. (See Secrist, "Statistics in Business," p. 77.)   The necessity of including goods with money in
considering true incomes was recognized by Pareto.

about lower incomes are known) not a straight line, it may be that these will remain useful indices, with the possible addition of a third coefficient of deviation. Meanwhile it must always be remembered that the significance of these coefficients in each community really belongs only to the theoretical straight-line community which the actual communities appear most closely to fit.

The selection of the two fixed points upon the actual curve, through which the fitted straight-line is projected (Chart II), is therefore of importance, and it is absolutely necessary that in any given set of coefficients the fixed points be the same for all communities. It is here proposed to set the fixed points where the curve intersects the ordinates of incomes of $10,000 and $100,000. Upon inspection, it will be seen that these parallel closely the general trend of the curve, without venturing too high in the curve where the number of observations (or reports) is small and consequently more subject to chance variations, and without coming into the zone of the more abrupt drop at the bottom of the curve. In certain investigations into the statistics of small parts of the population, it will be necessary to choose fixed points lower on the curve or nearer together, because the number of observations even at $100,000 becomes so small as to be subject to large chance variation; but when this is done the same points in the curve for the entire population should be inspected for their deviation, and the fitted straight-line of the small part of the population should be corrected by this deviation.*

Having fitted a straight line or norm, therefore, at two points, to the curve for any community, the two coefficients for the mathematical expression of this straight line are easily seen. The function of one of these, as already stated, must be quantitative, that is, it must fix the position of the straight line or norm, or of some standard point upon that norm, to the paper, graphically speaking. The function of the other is qualitative; it must define the slope or direction of the norm upon the paper. Let us first consider the latter.

If we define the slope as the tangent of the angle made by the norm and the ordinate or horizontal axis of the paper, we shall see that it may be expressed by the formula:

$$S = \frac{\log I - \log i}{\log n - \log N}$$

when I and i represent the larger and smaller income levels, respectively, at the two fixed points, and N and n represent the number of

---

* Some inconclusive study of the nature of the trend of the geometric ogive curve would seem to indicate that its changes in slope proceed in more regular intervals when the angle formed at the base is measured by circular measure. As S is the tangent of the angle, correction in the case described in the text should apparently be made after the angles have been from their tangents reduced into degrees.

persons at these levels.*   For use as an index or coefficient with popular connotation this factor of slope (*S*) is, of course, utterly unfitted, involving as it does logarithmic conceptions.   Examination will show that the anti-logarithm of *S* is no more than the ratio between two income levels so chosen that the numbers of persons at these two levels have a ratio of 1:10.   Likewise the anti-logarithm of 2 *S* would be the ratio between two income levels so chosen that the numbers of persons at these two levels have a ratio of 1:100.   In the latter case, if we divided the entire body of persons on the lower income level into groups of 100 persons each, and made the groups as representative† as
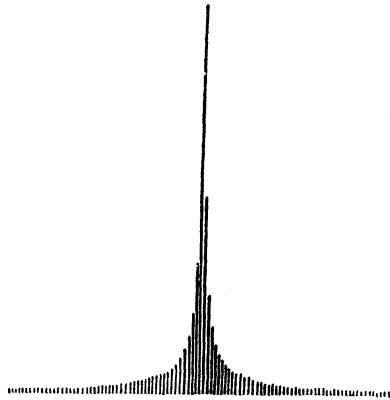

CHART III.   100 REPRESENTATIVE INCOMES

possible in the matter of high and low incomes, then in each such group there would be one person who is also in the higher income level.   Or we might say that if we segregated just 1 per cent of the population in such a way that every one in the 1 per cent group had an income greater than any one in the remainder, that is, if we segre-

---

* A short-cut method can be used if fixed points are always selected at two income levels, one of which is ten times as great as the other.   This has been done in the text and in the chart.   The result is that the phrase "log I—log i" or "log $\dfrac{I}{i}$" always resolves itself into unity and the formula can be written:

$$S = \frac{1}{\log n - \log N}.$$

† By representative is meant such sampling or selection in the arrangement of the groups, that each group will contain a unit whose ogive curve is identical in slant or slope, with the ogive curve of the entire community.   There would then be a considerable uniformity in the matter of the lowest income, but in many groups the highest income would be very great, since at any point along the ogive curve the number of persons plotted includes not only those with the given income, but also those with more than the given income.   Hence in these various groups the highest incomes would have many individuals with very large incomes; and the group here called representative is that group (by far the most common) whose wealthiest individual will have only the minimum of the high incomes.   The ratio is really between the maximum income exceeded by all of the single wealthiest individuals in the various groups, to the maximum income exceeded by all of the individuals in the same groups.

gated the wealthiest 1 per cent of the population, the anti-logarithm of 2 S would be the ratio of the minimum income in the wealthiest 1 per cent to the minimum income in the entire population. Here we have a concept which can be grasped by the layman. Roughly speaking, this index of distribution might be defined as "the ratio between the highest and the lowest incomes in a representative group of one hundred persons" (Chart III), or "the ratio of the income exceeded by the wealthiest 1 per cent of the population to the income exceeded by everybody."*

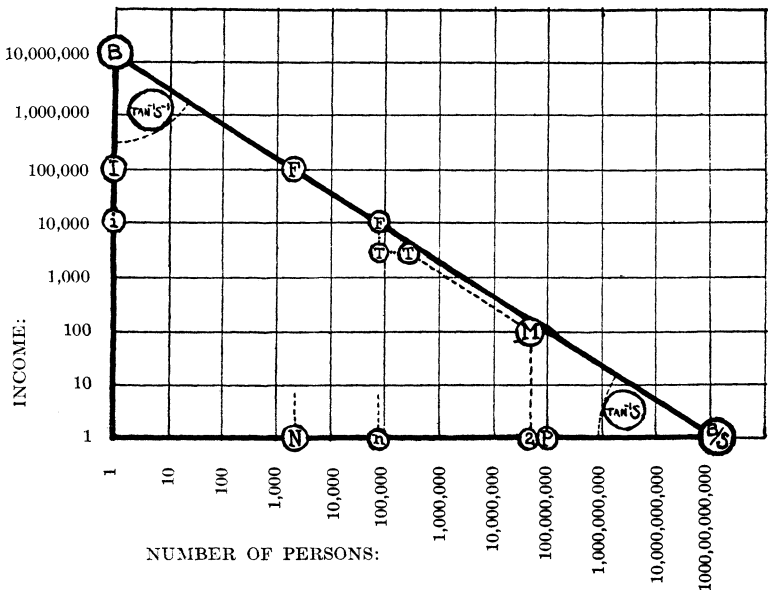$$\log D = 2\,S = \frac{2\,(\log I - \log i)}{\log n - \log N}$$

CHART IV. FULLY PROJECTED NORM

United States, 1914

Showing the fitted straight line or "norm" projected through the two fixed points (FF) at the intersections of the two known income levels (Ii) with the two numbers of persons (Nn) having more than these incomes. The line is projected to both axes, showing the bases (B and B/S) at its extremities. The computation of the median (M) is shown, one of the fixed points (F) being corrected for the number of persons represented by each tax-payer (TT) and a parallel to the norm being projected to the abscissa of one half (2) of the population (P), the intersection of the parallel line (dotted) and this abscissa (dotted) being at the ordinate of the median income (M).

* The trouble with this concept is that the minimum income of the entire population, or the income exceeded by everybody, is entirely theoretical, and takes no account of actual dispersion about the lower end of the norm. As a matter of fact, incomes do not cease suddenly at the lower extremity thereof, but before that point is reached, dispersion sets in and becomes quite marked, carrying a large number of incomes into lower levels, probably in the way of a normal probabilities curve.

Before proceeding to the other coefficient, the index of the quantity of wealth, it is interesting to note that from the formula for S we can derive the income level of any given portion of the population, or the part of the population on any given income level. That is to say, knowing S, i, and n in a community, and writing B for the constant $\log i + S \log n$ in that community, we may find the I for any N or the N for any I in that community, as follows:

$$\log I = B - S \log N \qquad\qquad \log N = \frac{B - \log I}{S}$$

Here we have the algebraic expression of the natural law of incomes stated above. Either formula may be traced graphically (Chart IV). The important point is that in either formula, while I and N are variables, there is a phrase which we have called B (in the second B/S) which is constant in the community. In the formula it serves as a base for the equation, in the chart it is seen to be the intersection of the norm with either axis.

In this case (B or B/S) we have the quantitative element in the formula, and it would serve if necessary as the quantitative coefficient. But examination will show that it is the logarithm of the number of dollars income of the wealthiest individual (B) or of the number of persons with incomes greater than one dollar (B/S), and that both numbers are fictitious and theoretical.* Hence not only because of its logarithmic conceptions, but also for its wholly fictitious nature, the base of the formula is not suitable as a coefficient for popular presentation. There is nothing to hinder our taking a point somewhere midway along the curve, however. Many such points suggest themselves, such as the intersection of the curve with the ordinate of $3,000. This would give us the number of taxpayers (in early years) as a quantitative coefficient, which we could read as a percentage of the entire population. Or we can adopt a certain abscissa, representing a certain

---

* The following definitions may be easily proved from the chart or the formula:

B is the logarithm of the theoretical income of the wealthiest individual in the community, measured in dollars of income, that is, the logarithm of the theoretical maximum income.

$\dfrac{B}{S}$ is the logarithm of the hypothetical and purely fictitious number of individuals in the community whose incomes exceed one dollar per year, and this number will greatly exceed the actual population of the community, for the reason that incomes do not exist as small as this amount.

S is the logarithm of the number of dollars income exceeded by one individual in every ten individuals, for every dollar of income exceeded by all of the ten; or in other words, the logarithm of the ratio of the two income levels exceeded by the wealthiest and the poorest individuals, respectively, in every ten persons.

$\dfrac{1}{S}$ is the logarithm of the number of individuals having more than any income, for every one individual having more than an income ten times as great, or in other words, the logarithm of the ratio of the two numbers of persons having more than any two income levels, one of which is ten times as great as the other.

given number of people, or a certain given percentage of the population. It is here proposed to adopt the abscissa of 50 per cent of the population. The constant for each community then will be the ordinate at which the curve intersects this abscissa. In short, it is proposed to use the median income in each community as the quantitative coefficient for that community. Here we have a concept which can be readily grasped by the layman, scarcely any concept would be easier; and yet for statistical purposes it is quite as good as either of the formulary bases, so long as we remember that it is, after all, at present only a coefficient, and does not actually represent dollars and cents of the median income. For at present it is the resultant of a projection of a fitted straight line through two distant points on the curve, which in themselves are only tax reports and subject both to a general upward correction for tax evasion and to a probable error which becomes magnified at the projected point.

It is not so easy to compute the median income from the formulary base $\left(B \text{ or } \dfrac{B}{S} \right)$ as it was to compute the index of distribution (D) from the factor of slope (S). In the first place, the total population of the community must be known. Where the community is the country or a state, it is not difficult, using estimates from the census. But in other cases, as where the community is an occupational rather than a territorial group, it is often impossible, and no median income can be given for such communities.* When we know the entire population, the question next arises as to how large a part of this population do the taxpayers represent. Obviously we cannot count the tax reports as representing only individuals; on the contrary, they are known to represent families. If we know the size of the average taxpayer's family, we can multiply the known number of persons (n) (representing families), on the known income level (i) to correct for the true number of persons (instead of families) on this level. (Though we have heretofore spoken of them as "persons" in this discussion, they are really families.) The resulting "I" will be the median family income. This assumes, however, that tax-paying families are the same in size as non-tax-paying families—a very questionable assumption. By further correcting the known income-level (i)—that is, dividing this income (I) by the number of persons in the taxpayer's family— we can get the income-level exceeded by the average individual in the

---

* Where no median income can be found for the entire population in any community or group, because the entire population is not known, it is sometimes useful to have at least a median taxpayer's income (family or individual), but such taxpayer's median, or median of incomes over a certain amount, cannot be used or corrected for use in comparison with medians of the whole population.

tax-paying classes. From this it is a simple matter to apply the formula and get the median individual income.*

In all cases, therefore, it is necessary to know the size of the average taxpayer's family before a median can be reached † Very definitely here we come to the point of estimates alone. Until 1917, no estimate whatever could be made from the income tax statistics. In that year for the first time personal exemption was allowed to taxpayers according to the number of persons in their immediate family who were supported by them. Hence from the total amount of personal exemption allowed, we may know the number of persons represented by the tax reports. The trouble is that this personal exemption was allowed only for heads of families, husband or wife, and dependent children under sixteen. In 1918 the law was altered to include all absolute dependents, so that not until the returns from 1918 are published, with the amount of personal exemption in that year, shall we have something like the true figure as to the size of the average family. In the meantime, however, the estimated medians, though palpably low, will retain for statistical purposes the full index value of the true medians.

When we come to applying this correction for families to earlier years, we will find that the number of wives, who, though living with their husbands, and part of the same family, make out separate reports from their husbands, varies greatly from year to year. And with any increasing odium which may attach to large incomes, the tendency thus to split up the family income formerly reported in its entirety by the

* The choice between a median family income and a median individual income depends finally upon whether the natural law of incomes (by this is meant the consistent natural forces operating to arrange human incomes into a giant geometric [or logarithmic] pyramid), applies fundamentally to bread-winners only or to all human beings. The considerations advanced below in the text would seem to place all human beings, whether dependent upon bread-winners or dependent upon employers and customers, alike under the operation of any natural law ordering their incomes.

† Two things militate against the use of a family median income. The first is the fact that we cannot entirely identify computed taxpayers' families (computed as in the text following) with the actual taxpayers' families, because in many families some children have independent or partially independent incomes, which prevents any exemption being allowed the head of the family for them. The second is the fact that we have not definite knowledge of the number or size of natural families in the United States, and particularly in the different states. The census lists (in census years) what it calls "artificial families," for the purpose of which neither blood nor marriage ties but common living and eating quarters are the basis of the family (residents at a boarding-house or hotel being counted as members of a single family), with the obviously questionable result that the urban and densely populated districts are credited with larger families than the rural and thinly populated districts.

Probably the most useful estimate of the size of the average American family can be obtained by dividing the total number of persons represented by the taxpayers, by the total number of tax-paying families. The latter (tax-paying families) is, of course, less than the number of tax-paying individuals, for the reason that many wives, though living with their husbands, make out separate tax reports, and it is still less than the number of adults in the tax-paying classes, for the reason that most wives make out no reports at all. According to this estimate, the natural family in 1917 comprised about 3.21 persons each, itemized as follows: "head of family or individuals living alone," average 1 person; "spouse," average .81 person; "dependent children under 18," average 2.40 persons. This statement must be qualified to the extent of its omission of "other dependents," explained in the text.

husband into parts, reported by both husband and wife, will increase. This process, be it remarked, probably had considerable to do with the decrease in the number of millionaire-income families reported from 1916 to 1917, over which so much publicity was stirred. There is, however, a very constant relation, which can hardly change appreciably from year to year, namely, the number of dependents represented per heads of families. Now the numbers of heads of families in the various years can be computed (Table III), and working from that relation we can proceed to estimates of the size of the average taxpayer's family, and from this to estimates of the median income either for families or individuals, remembering that all such estimates will be subject to uniform revision when the 1918 report is published. Method and comparative values will be the same, but the knowledge of the true size of the taxpayer's family will then merely inject a uniform upward correction.

The formula for the median incomes remains the same. In the belief that the median individual income is more valuable for statistical purposes than the median family income, the former is proposed as an index of the amount of wealth, the quantitative index of wealth. This belief is grounded in that fact that whether or not independent

TABLE III

COMPUTATION OF TAX-PAYING POPULATION. UNITED STATES, 1914–1917

NOTE: Figures in **bold-face** are taken from tax-reports, other figures are derived. The tax-paying population and the number of persons in this population per tax-payer (T) are subject to revision from later tax-reports not now available.

	(1917) (over $2000)
Joint returns of husbands and wives and separate returns of husbands	**1,450,723**
Separate returns of wives	**20,530**
Separate returns of husbands	20,530
Joint returns of husbands and wives	1,430,193
Wives not making returns, but represented by joint returns	1,430,193
Total adults making returns	**1,832,132**
Total adult population represented by returns	3,262,325
Total personal exemption claimed	**$3,773,079,800**
Adult personal exemption (@$1,000 each)	3,262,325,000
Dependent personal exemption	510,754,800
Total dependent population under 18 represented (@$200 each)	2,553,774
Total population represented by returns	5,816,099
Returns other than from families containing both husband and wife	**360,879**
Population represented by families containing husband and wife	5,455,220
Number of families containing both husband and wife	**1,450,723**
Persons per full family (containing both husband and wife)	3.76

	(1914)	(1915)	(1916)	(1917)
Returns married	**282,806**	**266,153**		
Wives separate returns	**3,985**	**4,819**		
Married couples	278,821	261,334	**355,107**	**1,450,723**
Persons per married couple	3.76	3.76	3.76	3.76
Population repres. by couple	1,048,250	980,000	1,335,000	5,455,220
Other returns	**74,709**	**70,499**	**74,294**	**360,879**
Total population represented	1,122,959	1,050,499	1,409,294	5,816,099
Total returns (over $2000)	**357,515**	**336,652**	**437,036**	**1,832,132**
T—Pop. repr. per tax-p	3.14	3.12	3.22	3.17

bookkeeping be maintained for each member of the household includ-
ing dependents, the truth is, strictly speaking, that every human
being enjoys the benefits of an income. The view may be taken that
dependence itself is really universal whether it be upon heads of
families or customers in business. This individual income figure will
perhaps correlate better with cost of living statistics, bank deposits,
mortality, and other sociological data.

The median individual income can be found from the formula:

$$\log M = \log \frac{i}{T} - S \left(\log \frac{P}{2\,Tn} \right)$$

where P and T represent respectively the entire population and the
ratio of the population in the tax-paying classes to the number of tax

TABLE IV

COMPUTATION OF THE INDICES. UNITED STATES, 1914

NOTE: The total population has been diminished by the tax-exempt population (Indians) to
get the net population (P). The number of persons per tax-payer (T) is subject to revision from
later tax-reports.

Item	Process	Sign	Anti-logarithm	Log.	Loglog.
Returns over $10,000................	Cumulation	n	80,788	4.9073	
Returns over 100,000................	Cumulation	N	2,348	3.3707	
Log n/N	Subtr. log.			1.5366	.1866
Log I/i...........................	Constant				.0000
S................................	Subtr. loglog	S		.6507	.8134−1
Distribution Index..................	Double log.	D	20.02	1.3014	
Returns over $10,000................		n		4.9073	
Persons per tax-payer...............	Computed	T	3.14	.4971	
2.................................		2		.3010	
2Tn..............................	Add. log.			5.7054	
Population total....................	Census		99,414,340		
Indians (tax-exempt)................	Census		335,998		
Population nett....................	Subtract	P	99,078,342	7.9960	
Log P/2Tn.......................	Subtr. log.			2.2906	.3599
S.............,..................		S			.8134−1
S(Log P/2Tn)....................	Add loglog.			1.4903	.1733
T................................		T		.4971	
Log T* S(log P/2Tn)	Add. log.			1.9874	
Income of $10,000.................	Constant	i		4.0000	
Index of Size.....................	Subtr. log.	M	102.95	2.0126	

reports. The median individual income can be used in the formulae
for finding unknown incomes by substituting for B, the phrase:
$\log M + \log \dfrac{P}{2}$, remembering that by so doing the reading of N has
been changed from families to individuals by the correction for T. In
the table (Table IV) the computation of both the coefficients is shown
in detail.*

* It is not deemed necessary to point out in the text that in computations of the median, the total
population must be corrected to the total population from which taxpayers are drawn. Thus tax
exempt groups, such as the American Indians, must be deducted.

3

It will be seen that the computed median for the whole United States in 1914 was little more than one hundred dollars. Professor King's estimate for 1910 was $800 per family, or about $175 per individual. Surely the median income did not fall so greatly during the interval. The explanation may lie in tax evasion and consequently incomplete reports which are the basis of the 1914 computed median. It may lie partly in the 1910 estimate. It may lie in the theory of the straight line, about which many statisticians are in doubt, and which can only be settled by a complete array of reliable figures in all income classes. It illustrates quite clearly, however, what has been already repeated, that the median computed from present tax reports cannot be taken as an actual measure of dollars and cents. Whatever value it may have at present is limited to use as an index figure in comparing different communities. Nor, with greatly changing efficiency in tax-collection methods, and some changes in tax-computing and tabulating methods, does it greatly interest us to know that the computed median for 1917 is 11 per cent higher than for 1914, for the entire United States. Its present use, even as a coefficient, is probably limited to comparisons of different parts of the country at the same time (Chart VI).

It is probably true that there are a number of real reasons for a lower actual median in the country than is generally supposed. For we must remember that the individual median is not the income which half the families exceed, nor half the bread-winners, nor half the adults; but it is the income which half of the entire population exceeds, and half falls short of. It takes into account, therefore, the actual income of children, babies, the sick, and public charges and dependents of all kinds. Many of these incomes are never thought of as individual incomes, but are covered in the bookkeeping of those upon whom the individuals are dependent. Nevertheless, it is an open question whether the incomes are not still as real in the last analysis as if separate accounts were maintained for each. But we must be prepared to meet with median income figures much lower than wages, possibly lower than cost-of-living standards. How much does the average child in a poor farming district live on, when in New York City, where probably twice as high a percentage of the population pay income taxes as in other parts of the country, the recent survey of the Board of Health shows 20 per cent of the school children suffering from malnutrition.

But whatever real causes there may be for a low median, there are artificial or statistical causes which definitely prevent the computed median (M) from being at present anything more than a statistical coefficient. Let us briefly review these causes. The computed median

is the mathematical result of projecting downward to 50 per cent of the population, a straight line which has been fitted to two points arbitrarily selected in the curve of the known income statistics which cover only 5 per cent of the population. Before this computed median can be taken as connoting a real sum of money, three hypotheses must be established: (a) That the curve of income statistics is really a straight line throughout all classes of society; (b) that the two points arbitrarily selected really lie in this straight line; and (c) that the number of dependents per taxpayers has been correctly computed. Until the straight line theory has been accepted in full the first of these hypotheses is in doubt. The second is openly untenable (the most that we can hope for being that the two points lie parallel to the true line) on account of tax evasion. The third admittedly awaits correction from figures compiled under the 1918 tax law and not yet available.*

Every care should, therefore, be taken not to use the present computed median as an actual sum in dollars and cents, but only as a statistical coefficient. For this purpose it is best used in terms of percentages of the average median or of some base reduced to 100 per cent (Chart VI). In such a use the variations in the computed medians, which would result from shifting the arbitrarily selected points through which the straight line is projected, lose importance, since if only the same points are used throughout all communities compared, the relative percentage of each will be about the same, regardless of the location of the points. And even as a coefficient alone, the computed median, or index of the amount of wealth, is subject to so large a probable error, owing to tax evasion, tax collection methods, and tax-computing and tabulating changes, that it certainly cannot yet be used for comparison between different years to much advantage and has only a questionable value in the comparison of different communities for the same year (Chart VI). It has been proposed here in the belief that in the course of time it will be possible to depend more fully upon the median income for quantitative wealth analysis, and that in the meantime there is need for some quantitative as well as a qualitative coefficient in the analysis of the present income statistics.

---

* The correction in the estimates of the taxpayer's family, which can be expected within another year, on the basis of fuller data as to dependents, will probably not increase the estimates from the present 2.4 persons to more than 3 persons per family   A common habit in American statistics has been to use the census "artificial family" figures of 4.5 persons per family (*cf.* footnote on p. 263). Let us assume, however, that the average American economic or natural family consists of 4.5 persons, in and out of tax-paying classes alike. Let us also assume that the tax collectors have reached only half of the wealth of the taxpayers; in other words, that tax evasion amounted to 100 per cent of the taxes reported. Both of these assumptions are in the nature of maximum possible errors. In the first case, a median income of $200 would be corrected to $175, and in the second case to $400, and in both cases at the same time, to $350.

The significance of this coefficient, in so far as it has any at present, may be illustrated as follows: Where one state shows a larger median than another, we may conclude that the majority of the population is wealthier in the first than in the second state. The median, therefore, indicates the degree of wealth or poverty of the great mass of the people in any community.* And so, according to this index, we can judge the condition of the masses in different states and communities, and in that sense can classify the communities themselves as "rich," average, and "poor" (Chart VI). It need not be added that to turn a coefficient of money wealth into a coefficient of real purchasing-power, the price-changes must also be considered.

It remains to make clear the significance of the other coefficient, the index of the distribution of wealth, which has been already computed. It is for sociological purposes far the more important of the two and, apart from the doubt attaching to the straight-line theory, and some smaller probable error due to tax evasion, is free of the weaknesses of the quantitative coefficient (the computed median). It has been roughly defined as the ratio of the highest and lowest incomes in each average or representative sample group of one hundred persons (or families) in the community. It denotes the geometric progression in which incomes increase in size as they decrease in number. Thus where the index of distribution is 60, there will be one individual with more than $60,000 a year for every hundred with more than $1,000; there will be one individual with more than $600,000 for every hundred with more than $10,000; and so on throughout the community's range of incomes. An increase in the index will mean a higher income for the one individual; a lower index, a lower income for him. So a dispropor-tionate rise in the fortunes of the few or a disproportionate fall in the fortunes of the many, will be reflected in a higher index of distribution; and vice versa, a disproportionate fall in the fortunes of the few or rise in the fortunes of the many will be reflected in a lower index of distribution.

It is not the actual wealth of the people measured in dollars and cents, nor even in terms of price changes, which determines their buy-ing power in the long run so much as the relation of their wealth to the wealth of others in the same community. Prices themselves appear

---

* Were it possible to obtain the average per capita income, the latter might be advanced in place of the median as a coefficient of wealth. Habit has already gone far to establish the per capita wealth (not merely income) as such a coefficient. But it is submitted that while the simplicity of the computa-tion of the mathematical (or arithmetical) average has much to recommend it, the latter figure has not an iota of the significance which attaches to the median in this connection. Commonly the median has not been obtainable, because the distribution of incomes was lacking. When it can even be estimated, it is believed that its wealth of human significance places it easily above the arithmetical average in usefulness.

to be affected, not alone by the general wealth but also by this compara-
tive or relative distribution. The trend of modern economics is placing
ever greater importance upon this comparative consuming power of
the different members of the community, and attributing to it an
influence upon the general level of prices and indeed upon the entire
trend of productivity in the community. Into the effects of the injec-
tion of many relatively large incomes into a community, upon prices
in that community, and upon the shifting of production objectives from
the so-called necessities to the so-called luxuries, it is not the province
of this discussion to go; but it must be seen that it is precisely the ex-
tent of such injection of relatively large incomes of which the index of
distribution is a measure.

It may even be argued that the index of distribution is the measure
of the good or bad health, economically speaking, of income and wealth
conditions in a community; that its variations reveal more truly than
other indices the actual changes of distribution in all that relates to
incomes and earning power, and to a considerable degree serve to mark
the even less tangible forces of human happiness, citizenship, and both
producing and consuming powers of the individual. We tread here
upon ground so delicate that any heavy pressure of the point will
break through completely, and the utmost that can be defended in
such a general interpretation is that the index of distribution, *ceteris
equibus*, records a very vital and sensitive factor in the shifting condi-
tions of human life.

To be specific, let us assume two extreme conditions of society, the
one where wealth is equally enjoyed by all, and the other where one
individual owns and enjoys all the wealth of the community. Such
conditions are purely hypothetical, but for lack of a better phraseology
the one may be described as communistic and the other as economically
monocratic. In these communities, the index of distribution would
be, in the first case, unity, in the second infinity. Between these two
extremes we may take as a mid-way community one in which the
number and sizes of incomes exceeded would vary inversely,* and the
index of distribution would be 100.

What is the significance of this mid-way community? Perhaps it
can be illustrated briefly by assuming that we wish to sell in this com-
munity an article which every one who can will purchase, but will
purchase only once. In the mid-way community, it makes no differ-
ence what price we set, our gross sales will amount to the same sum.

* In such a community the second richest person would have an income of one-half the income of
the richest person, the third richest person would have one-third, the fourth would have one-fourth,
and so on down to the one-hundred-millionth person, who would have but one one-hundredth-millionth
of the income of the richest individual.

For in this community, for each and every person having a hundred dollars to spend, there are five persons having $20 each, ten persons with $10 each, and one hundred who have one dollar to spend. Whether our price be one dollar or a million dollars, we shall have the same gross receipts from sales. But as the index falls below 100, our potential market is greater if a lower price be set; when the index is greater than 100, our potential market is larger if the price is raised. Of course, in actual cases many other factors would enter into buying motives and price fixing. The case is here put only to illustrate a balance in the mid-way community.

Adopting this mid-way community whose index is 100, therefore, as a turning point between two conditions in which the wealth of the many poorer or the few richer predominates, we find ourselves at a loss for words with which to describe the significance of the two conditions. It may be observed that in a more democratic distribution of wealth the index will be lower, while in a more plutocratic distribution the index will be higher, and that to either trend the response of the index is immediate. Exception may be taken to the use of the over-worked words "democratic" and "plutocratic" in this connection, but the present writer can find no others which carry the connotation. Without wishing to abuse the meaning of these words, they are offered as the best translation of the two vitally significant conditions described. Where the index is less than one hundred, the community would then be described as economically more "democratic" than "plutocratic"; where the index is more than one hundred, the community would be called economically more "plutocratic" than "democratic"; and in comparing different communities by the index of distribution of wealth, we could classify them—always remembering the inadequacy of the words—as economically "democratic," average, or "plutocratic."*

The trend of the United States in the last four years in which statistics are available, as shown by this index of the distribution of wealth, is extremely interesting (Table V)   It will be seen that throughout the early period of the European war the rise of the fortunes of the few was rapid and the conditions of wealth distribution were leaving a point of great economic "democracy" and rapidly approaching a point of economic "plutocracy." We see that in 1917, upon the entrance of this country into the war, this course was halted and a slight return toward "democracy" set in. The cause of this reversal of trend may

---

* For certain statistical purposes the weakness of any arithmetical relative index applies to the index of distribution, in that changes of similar character are more exaggerated in amount in the large indices than in the small ones, precisely as happens in relative price indices which are measured arithmetically. If the logarithm of the indices (of distribution) be used, the changes will be marked more evenly. Or if the index of distribution is plotted on logarithmic paper the same result is reached. Either step is virtually the adoption of the factor of slope (S) from which the index of distribution (D) was computed.

perhaps be found in the governmental interference and price-fixing which accompanied the conduct of the war. If so, the measure of government control experienced by this country was extremely bene-

TABLE V

INDEX OF DISTRIBUTION OF WEALTH. UNITED STATES, 1914–1917

Year	Index (D)
1914	20.02
1915	29.31
1916	38.26
1917	27.75

ficial to the great majority of the population for at least the time being. The movement of the index of wealth distribution in later years as figures become available from the tax collector's office, will be of great interest in showing how permanent or temporary was this trend away from "economic plutocracy."

An interesting comparison can also be made with the countries included in Pareto's survey. In the table (Table VI) the conditions noted by him are set forth, after translation into the index here

TABLE VI

DISTRIBUTION OF WEALTH ABROAD. (PARETO, "COURS D'ECONOMIE POLITIQUE")

Country	Province	City	Urban or Rural	Year	Index of Distribution (D)
France...................	....	Paris (rents)	Urban		18.8
England.................	....	....	Total	1843	21.5
				1880	30.3
Germany.................	....	Augsburg	Urban	1471	25.0
				1498	22.9
				1512	38.8
				1526	59.4
	Saxony	....	Total	1880	18.4
				1886	21.1
	Prussia	....	Total	1852	11.4
				1876	14.6
				1881	14.3
				1886	15.5
				1890	17.8
				1894	17.8
Italy....................	Anconia, Arrezzo, Parma and Pisa		Urban		32.7
	Florence		Urban		26.2
	Perugia	....	Urban		15.3
			Rural		28.8
	....	Misc. cities	Urban		24.0
Switzerland...............	Basle		?	1887	41.0
Peru.....................	....	....	Total	ca. 1800	13.1
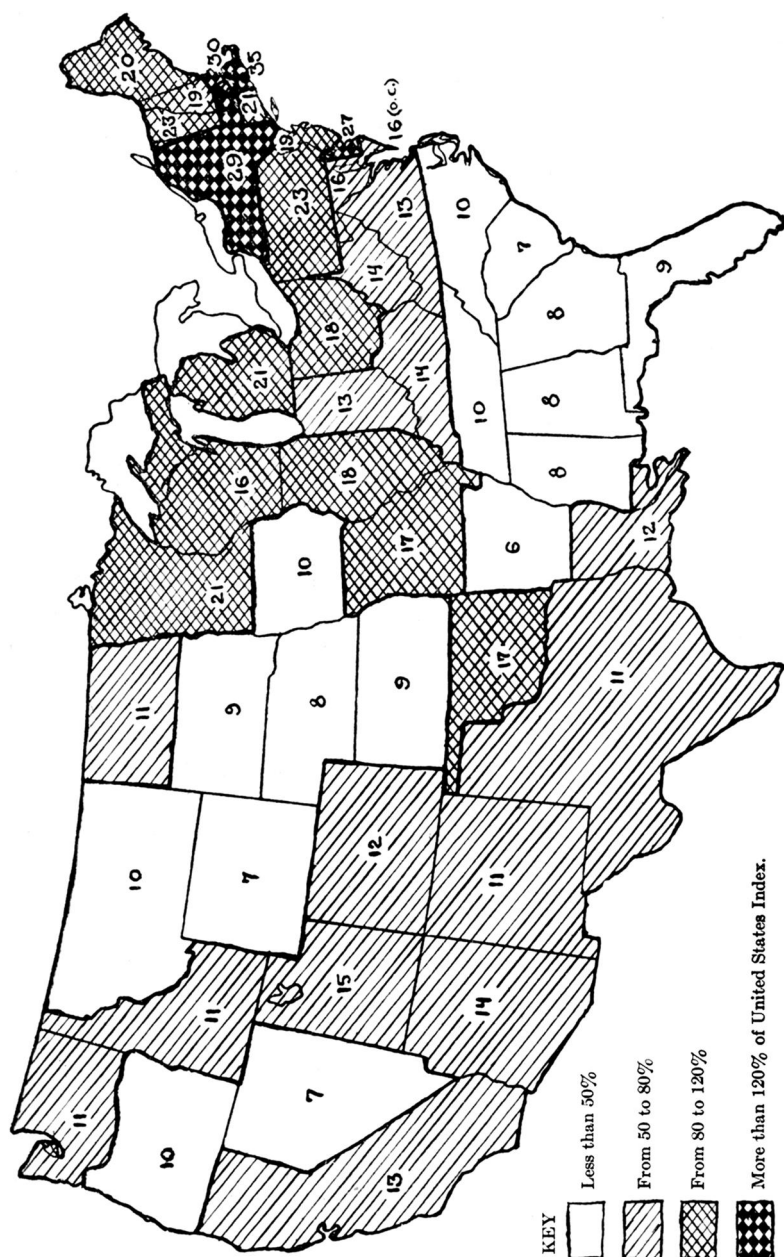
CHART V. DISTRIBUTION OF WEALTH
United States, 1914
(United States Index—20)

KEY

Less than 50%

From 50 to 80%

From 80 to 120%

More than 120% of United States Index.

adopted.*   The history of the city of Augsburg in the fifteenth and sixteenth centuries is extremely significant of social conditions in a decadent community drifting from democracy to plutocracy. The same trend, less marked, is evident in the Prussian figures.

Another interesting use of the index may be made from the detailed statistics of income prepared by the government in 1916 (Table VII)

TABLE VII

DISTRIBUTION OF WEALTH BY OCCUPATIONS. UNITED STATES, 1916

Occupations	Index	Occupations	Index
Total (All groups)...................	38.3	Business Group (*continued*)	
Professional Group..................	15.0	Production......................	61.3
Professions proper................	14.1	Agriculture.....................	18.8
Military.......................	7.6	Lumber-owners.................	30.1
Public service—civil..............	11.5	Mine-owners and operators........	217.5
Law.........................	22.9	Manufacturers..................	93.6
Clergy.......................	9.2	Distribution.....................	29.6
Medical......................	7.7	Merchants and dealers............	34.3
Teaching.....................	8.3	Travelling salesmen..............	6.9
Fine arts......................	18.5	Catering........................	14.8
Writing......................	17.7	Hotel owners, etc................	13.1
Stage and music..............	31.4	Saloon keepers..................	8.5
Art and sculpture.............	13.2	Sportsmen, turfmen, etc..........	9.3
Architects....................	15.3	Theater owners.................	39.0
Technical......................	14.7	Other business..................	35.2
Engineering..................	16.9	Business not stated..............	56.2
Accounting.	10.1	Financial Group....................	83.2
Other professions...............	20.6	Bankers.......................	89.1
Profession not stated.............	17.6	Insurance brokers...............	22.6
Business Group.....................	31.2	Insurance agents, etc............	11.8
General.......................	18.9	Real estate brokers..............	19.3
Corporation officials.............	29.6	Stock & bond brokers.............	595.0
Other office employees...........	18.3	Other brokers...................	76.3
Laborers......................	12.3	Investors and speculators..........	95.6

in which occupational analysis was made. It will surprise few to find the professions so very much more " democratic," economically speaking, than the business or capitalistic occupations.

A geographical analysis of the distribution of wealth cannot be so accurately made from the reports, for reasons connected with the method of reporting,† but in spite of its greater margin of error, such

* Pareto's measure was the tangent of the angle formed by the fitted straight line to the axis on which the incomes were plotted. As explained in the footnote on p. 253, this is the reciprocal of the factor of slope (*S*) here used. So that the index of distribution quoted in the table is simply the anti-logarithm of twice the reciprocal of his measures.

† "Income reported by an individual or corporation in one state may have been derived from sources in other states." (Statistics of Income, 1916, p. 12.)   Hence the figures show geographical distribution of receipt of incomes, but not of source of incomes. A subsequent statement, that "for the collection of tax at the source the collecting agent filed the return in the state in which the agent was located, regardless of the residence of the person earning this income," applies only to corporate returns, not to personal returns, since payments of tax at source are credited to individuals wherever they reside.

The year 1916 is exceptional in income statistics, and really to be compared with other years, for the reason of a change in reporting adopted in that year only. This change, introducing a statistical difficulty which can hardly be overcome, is as follows: If A and B, married—husband and wife—filed separate reports, A for $10,000 and B for $5,000, they were entered in the final report in 1914, 1915, and 1917 as two returns of $10,000 and $5,000 each, but in 1916 they were entered as one return of $15,000, and in the occupational analysis of 1916 they were entered as two returns of $15,000 each.
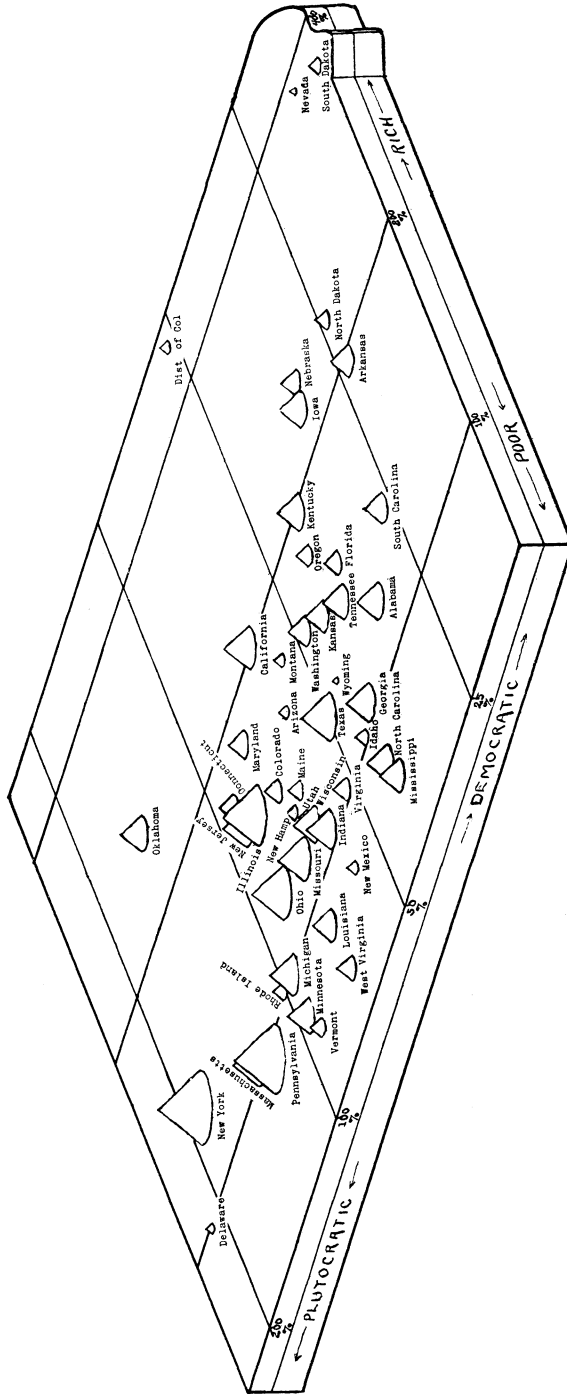
CHART VI. WEALTH CONDITIONS IN VARIOUS STATES

United States, 1916

Showing each state as a cone, of which the area of the base represents the population of the state, and the position upon the two axes, the two indices of wealth (index of size and index of distribution) of the state as compared to the two indices for the entire United States (U. S. = 100% on both axes). It should be observed that the indices for Nevada and South Dakota are based upon such scant returns as to make them quite unreliable, and their extreme position has little significance. The difference between urban populations and total urban and rural populations is perhaps illustrated by the position of the District of Columbia.

an analysis is of great interest (Chart V). Note for example that the entire south, where the median income is low, and the wheat belt, where it is high, both show a very "democratic" structure. Delaware, Rhode Island, Massachusetts, and New York are extremely "plutocratic." Iowa, a farming state of very high median income, where the number of automobiles per capita is largest, has a low index of distribution, while Oklahoma, also an agricultural state, but rich in mineral resources, has a high index.

Instances of the use of this index could be multiplied greatly, as also of the use of the quantitative index, but it is believed that the foregoing sufficiently illustrate the method. In conclusion, we may say that the two indices or coefficients here suggested for the measurement of income conditions, rest upon the following steps:

(a) Statistics of income are cumulated downward to show the number of persons having more than each amount of income.

(b) The cumulated statistics are plotted upon double-logarithmic paper and a straight line is fitted to their curve at two points arbitrarily selected.

(c) The two points are selected arbitrarily at the intersection of the curve and the ordinates of the incomes of ten thousand and one hundred thousand dollars.

(d) At the intersection of the projected straight line and the abscissa of half of the population, after the necessary corrections for the size of tax-paying families, the indicated median individual income is read from the ordinate at this intersection and is taken as an index or coefficient of the size of incomes in the community.

(e) The tangent of the angle made by the fitted straight line and the base or horizontal axis of the paper, upon which the number of persons has been projected, is taken as one-half of the logarithm of the index or coefficient of the distribution of wealth.

By the use of these two indices we can expect to classify communities as relatively "rich" or "poor" and (economically or financially) "democratic" or "plutocratic" (always understanding these expressions in the limited sense above detailed). We can look for communities of four classes, namely: "rich and democratic," "rich and plutocratic," "poor and democratic," and "poor and plutocratic"; in short we may have every combination of the two conditions in every degree (Chart VI).*   And the great interest will not lie so much in the deter-

* A useful study would consist of the collection of the two indices for a large number of groups of the population with the idea of finding any degree of correlation which may exist between the two. *Prima facie* there would appear to be none, and it is, of course, true that any statistical correlation based upon medians which are not fairly reliable may be attributed to statistical errors (for the index of distribution enters into the projection downward to the median, and is, therefore, a component part of the

mination of these conditions at any given point of time, as in the discovery of the nature of the changes in these conditions in the community or communities, from time to time. For society is rarely static; it is normally dynamic, and the chief problem of interest is where, when, and how it is moving.

median). But some preliminary examination seems to indicate a rather large degree of correlation, whereby the larger the index of distribution the smaller the median income. This will be seen in the chart (Chart VI) where most of the states are grouped in a broad zone running diagonally across the board. If such a correlation can be established, the significance is very great, as it will mean a sweeping revision of our present basis of judging prosperity. That is to say, if it can be proved that, using the terminology of the text, "plutocratic" communities are normally "poor" and "democratic" communities are normally "rich," it means that as the volume of visible wealth owned by the very rich members of the community grows larger and larger, the condition of the great mass of people in that community is not improving, but is deteriorating, and great wealth will in itself become one of the signs of such deterioration.